Journal of the Royal Society Interface 2014 vol.11 N101

Universals versus historical contingencies in lexical evolution

Bochkarev V., Solovyev V., Wichmann S.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2014 The Author(s) Published by the Royal Society. All rights reserved. The frequency with which we use different words changes all the time, and every so often, a new lexical item is invented or another one ceases to be used. Beyond a small sample of lexical items whose properties are well studied, little is known about the dynamics of lexical evolution. How do the lexical inventories of languages, viewed as entire systems, evolve? Is the rate of evolution of the lexicon contingent upon historical factors or is it driven by regularities, perhaps to do with universals of cognition and social interaction?We address these questions using the Google Books N-Gram Corpus as a source of data and relative entropy as a measure of changes in the frequency distributions of words. It turns out that there are both universals and historical contingencies at work. Across several languages, we observe similar rates of change, but only at timescales of at least around five decades. At shorter timescales, the rate of change is highly variable and differs between languages. Major societal transformations as well as catastrophic events such as wars lead to increased change in frequency distributions, whereas stability in society has a dampening effect on lexical evolution.

http://dx.doi.org/10.1098/rsif.2014.0841

Keywords

Google Books N-grams, Kullback-Leibler divergence, Language dynamics, Lexical change